AD_____

Award Number: DAMD17-99-1-9323

TITLE: Antigene Strategy in Breast Cancer Therapy: Rationales for Direct Targeting of erbB2/Her2 DNA with Polyamides

PRINCIPAL INVESTIGATOR: Vsevolod Y. Katritch, Ph.D.

CONTRACTING ORGANIZATION: The Scripps Research Institute
La Jolla, California 92037

REPORT DATE: September 2001

TYPE OF REPORT: Annual Summary

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	September 2001	Annual Summary (1 Sep 00 - 31 Aug 01)

4. TITLE AND SUBTITLE
Antigene Strategy in Breast Cancer Therapy: Rationales for Direct Targeting of erbB2/Her2 DNA with Polyamides

5. FUNDING NUMBERS
DAMD17-99-1-9323

6. AUTHOR(S)
Vsevolod Y. Katritch, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
The Scripps Research Institute
La Jolla, California 92037

E-Mail: sevak@scripps.edu

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 Words)*

The major goal of this project is to design and computationally evaluate the most potent Pyrrole-Imidazole containing polyamide inhibitors of erbB2/Her2 transcription. In the first year of the project we used an original algorithm to identify the most suitable sequences within erbB2 promoter DNA to target with the polyamides. This year we focused our efforts on generating polyamides with high affinity and specificity to the target DNA sequences. Thus, we have developed an improved and extended version of our algorithm to build 3-Dimentional molecular models of polyamides bound into the minor groove of DNA. In this algorithm, a standard B-DNA model is used as initial conformation for double stranded DNA, and the polyamide chain is placed in the minor groove according to the specified polyamide-DNA pairing rules. Geometries of these initial conformations are energy optimized under special distance constraints, derived from the available polyamide-DNA X-ray structures. The improved approach provides reliable models for longer complexes and allows to include new elements into polyamide design. In addition we developed a fast method to introduce "mismatches" into the DNA sequence of the model and evaluate their effect on affinity of the complex. Systematic study of DNA-polyamide complex with such mutations can be used to predict binding specificity.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Breast Cancer	12
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

Table of Contents

Cover.....

SF 298.....

Table of Contents.....

Introduction..... 3

Body..... 3

Key Research Accomplishments..... 7

Reportable Outcomes..... 7

References..... 8

Appendices..... 9

Introduction

Pyrrole-Imidazole (Py-Im) containing polyamides is a new class of DNA minor groove binding molecules that use a set of well-characterized pairing rules to recognize dsDNA sequences with high affinity and sequence specificity, comparable to affinity and specificity of gene transcription factors (Kielkopf *et al.*, 1998; Wemmer & Dervan, 1997; White *et al.*, 1997; White *et al.*, 1998). In addition to Pyrrole and Immidazole rings and their modifications, polyamide chains may contain other "residues" that improve binding specificity (Wang *et al.*, 2001) or prevent binding of activator proteins(Bremer *et al.*, 1998; Bremer *et al.*, 2001).

Our project is designed to target the erbB2/Her2 promoter region with polyamides in order to disrupt formation of the transcription complex and thus inhibit production of this important oncogene. Earlier in the project we identified the optimal sites within erbB2 DNA promoter sequence which 1) interfere with binding of transcription factors 2) have maximum genome specificity 3) are suitable for polyamide design. In this report we describe the latest results of 3-D molecular modeling aimed to find optimal polyamide binders for these target sequences.

Body

Task1: Optimization of target sequences in the Her2/erbB-2 promoter.

This part of the project was described in the previous annual report. We found that the region around the TATAA box, which is very important for regulation of gene activity, has very poor specificity in the human genome(Chiang *et al.*, 2000). On the other hand, we discovered sequences containing 13bp fragments with almost unique whole-genome specificity, also overlapping with one or more erbB2 activation sites. As a result, the following erbB2 promoter sequences have been chosen as optimal targets for polyamide design:

No	DNA Sequence	Regulatory elements possibly involved
1	**AG**TTGCCGACTC**CCAG**	GC box element and Thing1/E47 heterodimer
2	CTTCGTTGGAATGCA**G**	c-Myb
3	**GAG**CGCGCTTGCTC**CC**	COMP1 and CCAAT box
4	**AGG**AGGGCTGCTT**GAG**	VDR/RXR heterodimer, AP2, c-Ets-1

Task 2: Overall design and evaluation of complimentary polyamides.

Preliminary design of polyamides matching target DNA sequences was largely accomplished last year, resulting in a database of polyamide structures with different

combinations of rings and aliphatic substitutes.

We revisited this task recently, after a new important polyamide residue, *N*-diaminoalkylpyrrole, was added to the polyamide design repertoire (Bremer et al., 2001). Polyamides with diaminoalkyl "positive patch" not only allow reliable inhibition of transcription factors with exclusive major groove binding, e.g. bZIP proteins, but improve affinity and specificity of DNA recognition. Thus, using alkylpyrrole positive patch and C-terminal N-methylamide as a "tail" can improve polyamide gene inhibitors in many cases (Figure 1).

We designed and optimized new *N*-diaminoalkylpyrrole, *N*-diaminoalkylimidazol and N-methylamide residues, and incorporated them into the library of polyamide elements.

Task 3: **Detailed modeling and selection of candidate structures**

a. Test and adjust the ICM global minimization procedure with published polyamide-DNA complexes.

We have developed an improved automated procedure to generate 3-D molecular models of polyamide–DNA complexes, making modeling more reliable for longer complexes with new design elements.

The first improvement deals with the choice of starting configurations of the complex and polyamide placement. The new algorithm uses standard B-DNA as initial conformation, and places the polyamide chain into the DNA minor groove according to the specified polyamide-DNA pairing rules. Only then the special distance constraints, provided by the available polyamide-DNA X-ray structures are employed in the energy optimization of the complex. These modifications allows to avoid strong deviations from B-DNA structure in the initial steps of the procedure and provide much better convergence for energy minimizations,

The other improvement takes advantage of the new internal coordinate force field (ICFF) developed in the lab (the paper on ICFF is submitted to the Journal of Computational Chemistry). The ICFF is automatically generated from a "source" Cartesian force field (such as MMFF94s or Amber) with an algorithm that "projects" Cartesian parameters into the torsion coordinate space. Implicit flexibility, naturally incorporated into the internal coordinate parameters, is critical to the accuracy of the internal coordinates model with fixed covalent geometry. Essential also is the ability of ICFF method to generate fixed covalent geometries for new chemical structures, using Cartesian geometry minimization with the source force field. This feature facilitates inclusion of new elements into our custom polyamide residue library, producing residue geometries compatible with the new force field.

Prediction accuracy of the new algorithm with ICFF geometries and energy functions substantially improved compared to the previous version with ECEPP torsion potential, reducing geometry RMSD from ~1.2 Å to just ~0.9 Å in our standard test with available PDB entries (365d and 334d). Binding free energy estimations with the new algorithm also improved from 1.7 kcal to 1.3 kcal RMSD (Figure 2.)

b. Build all-atom models for DNA complexes with newly designed polyamides.

The automated procedure for polyamide design was programmed with ICM molecular

modeling package, which takes DNA sequences and coded polyamide sequences as input, and produces energy optimized complexes in the output. An example of the program output is shown in Figure 3.

The program reads the input sequence where each DNA and polyamide "residue" is represented with one letter or symbol. Double stranded DNA is built in a standard energy optimized B-form by an original ICM script. A polyamide chain of specific sequence (or two chains in case of overlapping hairpin topology) is built from the library of residues. The pairing between polyamide residues and DNA residues is assigned according to the input. One or more X-ray templates are then superimposed with the DNA structure to cover the polyamide binding site, and the polyamide atoms are "tethered" to the corresponding polyamide atoms in the templates.

Tight binding of polyamides in the DNA minor groove and the modular nature of the pairing between the molecules suggest special approach to energy minimization of the complex. We apply so-called ICM "regularization" procedure to minimize both length of the "tethers" and the conformational energy of the object. Regularization procedure goes through several iteration steps, using different weight ratio for conformational energy and "tether tension" energy at each minimization step. The weight of the tethers in the energy function gradually decreases throughout the regularization procedure, making the final solution virtually independent on the tethers. Minimizations, performed in torsion coordinates, not only guarantee fast convergence of this procedure, but also prevent severe deformations in covalent geometry due to the tether tension in the initial steps of the procedure. Spatial positions of the templates are readjusted in the course of the regularization procedure to allow large-scale movement of DNA backbone. This annealing-like algorithm is designed to generate low-energy structures with high local similarity to the templates.

For each of the four selected 16-bp DNA targets, we generated more than 100 polyamide "perfect match" complexes with 12-bp DNA recognition sites, which differ in positions of 5-member rings in the sequence or in overall topology. We use several criteria to check the quality of the models built. First, we check the length of hydrogen bond contacts between polyamide and DNA residues, which are expected by the pairing rules. For the best models we found 93% of the of the 34 hydrogen bonds within 2.5 Å lengths (measured as hydrogen to heavy atom distance), while on the average about 89% of the H-bonds satisfy this criteria for the "perfect match" models. Second, we check the tethers between the model and the template, and found that the average length of the tethers is about 0.5 Å and usually do not exceed 1.5 Å. Finally, we performed ten independent runs with single mismatches in the polyamide sequences and found the consistent increase in the complex conformational energy compared to the perfect match case.

c. Calculate global minimum conformations for each complex and evaluate polyamide-DNA binding energy.

The annealing procedure, employed in the global energy optimization of the complex is described above. We performed a separate study with three polyamide-DNA complexes to assess global convergence of energy optimizations in our special case. For each model we used 20 independent runs of the procedure with different annealing schedules. In all the three cases we found slight variability in the results of different runs,

with the average conformational energy RMSD ~0.7 kcal and geometry RMSD~1.3 Å. Such conformational variability is expected in the polyamide-DNA complexes, and has to be taken into account by averaging results over several independent runs.

Much more flexible aminoalkyl and C-terminal methylamide moieties of polyamides were treated separately with the ICM Monte Carlo global optimization method to allow large-scale changes in their conformations. ICM allows freezing of the variables in the rest of the complex, which makes exhaustive Monte Carlo search in the flexible parts of the molecule possible on a reasonable time scale. We found this Monte-Carlo search critical to avoid local minima trapping of the flexible parts of the polyamide molecule.

Polyamide-DNA binding energy for a given conformation of the complex was predicted as a sum of hydrogen bonding, van der Waals and electrostatic interactions energies between polyamide and DNA, combined with different weights (1., 0.43 and 0.75 respectively). This binding energy formula was previously found to be optimal by calibration with available experimental results (see Figure2.) For each polyamide-DNA complex, the binding energy was calculated as an average of binding energies of five independently minimized conformations. Binding energy results for some of the suggested polyamide binders are presented in Table 1. We plan to synthesize and test affinity of these polyamides in collaboration with Prof. David Wemmer group at UC Berkeley.

Key Research Accomplishments

- We have included new aminoalkyl-modified residues in the polyamide residue library, improving both affinity and inhibitory effect of the designed polyamides
- We have upgraded the modeling algorithm, making feasible reliable calculations for longer complexes with new design topologies
- Using the automated procedure, we have generated more than 400 "perfect match" polyamides, targeting four most important activation sites in the erbB2/Her2 promoter sequence.
- We have written a program and used it to generate all-atom models for all the 400 polyamide-DNA complexes, based on the known pattern of polyamide-DNA recognition and on the global geometry optimization
- We have predicted binging energy of these polyamides and selected most potent ones for further experimental studies.

Reportable outcomes

- Meeting Presentation and Abstract:
Bernhard H. Geierstanger, Colin J. Loweth, Vsevold Katritch, Ruben Abagyan, Peter G. Schultz & David E. Wemmer (2001).
NOE distance constraints and structural modeling of a ten-ring hairpin complex with DNA. *Frontiers of NMR and Molecular Biology Meeting,* Keystone, CO.

- Articles:

Vsevolod Katritch, Maxim Totrov and Ruben Abagyan (2001). ICFF: A new method to incorporate implicit flexibility into an internal coordinate force field. Submitted to *J. of Comp. Chem.*

The modularity of DNA recognition by polyamide molecules persists for a ten-ring hairpin in complex with an eight base pair binding site.
Bernhard H. Geierstanger, Colin J. Loweth, Vsevold Katritch, Ruben Abagyan, Peter G. Schultz & David E. Wemmer. (2001) Submitted to *J. of Am. Chem. Soc.*

References:

Bremer, R. E., Baird, E. E. & Dervan, P. B. (1998). Inhibition of major-groove-binding proteins by pyrrole-imidazole polyamides with an Arg-Pro-Arg positive patch. *Science* **282**(5386), 111-5.

Bremer, R. E., Wurtz, N. R., Szewczyk, J. W. & Dervan, P. B. (2001). Inhibition of major groove DNA binding bZIP proteins by positive patch polyamides. *Bioorg Med Chem* **9**(8), 2093-103.

Chiang, S. Y., Burli, R. W., Benz, C. C., Gawron, L., Scott, G. K., Dervan, P. B. & Beerman, T. A. (2000). Targeting the ets binding site of the HER2/neu promoter with pyrrole- imidazole polyamides. *J Biol Chem* **275**(32), 24246-54.

Kielkopf, C. L., Baird, E. E., Dervan, P. B. & Rees, D. C. (1998). Structural basis for G.C recognition in the DNA minor groove. *Nature* **391**(6666), 468-71.

Wang, C. C., Ellervik, U. & Dervan, P. B. (2001). Expanding the recognition of the minor groove of DNA by incorporation of beta-alanine in hairpin polyamides. *Bioorg Med Chem* **9**(3), 653-7.

Wemmer, D. E. & Dervan, P. B. (1997). Targeting the minor groove of DNA. *Chem Biol* **4**(8), 569-78.

White, S., Baird, E. E. & Dervan, P. B. (1997). On the pairing rules for recognition in the minor groove of DNA by pyrrole-imidazole polyamides. *J Mol Biol* **274**(4), 439-45.

White, S., Szewczyk, J. W., Turner, J. M., Baird, E. E. & Dervan, P. B. (1998). Recognition of the four Watson-Crick base pairs in the DNA minor groove by synthetic ligands. *Chem Biol* **5**(3), 119-33.

Input sequence	Predicted binding energy, kcal
GAGCGCGCTTGCTC**CC** **#cyclic** IPIbIPbPIR g g PIPbPIbPPI **CTC**GCGCGAACGAG**CC**	-18.3 ± 1.5
GAGCGCGCTTGCTC**CC** PIbIPbPIPR g mbIPbPIbPPIP **CTC**GCGCGAACGAG**CC**	-18.2 ± 1.7
GAGCGCGCTTGCTC**CC** PIbIPbRIPP g mbIPbPIbPPIP **CTC**GCGCGAACGAG**CC**	-18.0 ± 1.4
GAGCGCGCTTGCTC**CC** IPIbIPbPIR g mbPIPbPIbPPI **CTC**GCGCGAACGAG**CC**	-17.3 ± 2.5
GAGCGCGCTTGCTC**CC** PIPbPPbIPR g mbIPIbIPbPIP **CTC**GCGCGAACGAG**CC**	-16.9 ± 1.7

Table 1. Top five suggested polyamide binders. Accuracy of the energy predictions was accessed by five independent annealing minimizations.

One-letter codes for polyamide residues are: P- pyrrole, I- Imidazole, H- hydroxypyrrole, b- β-alanine, g- γ-linker, K- diaminoalkylpyrrole, R- diaminoalkylimidazole.

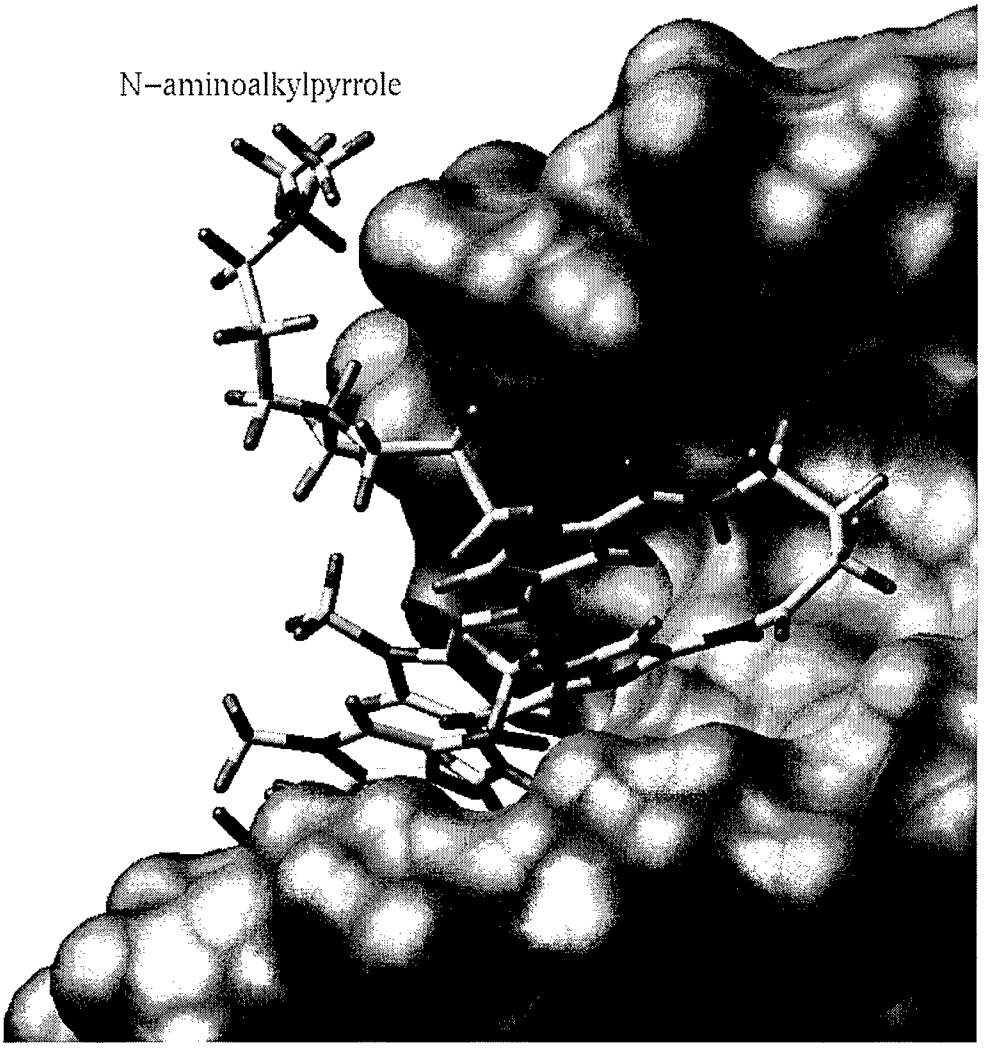
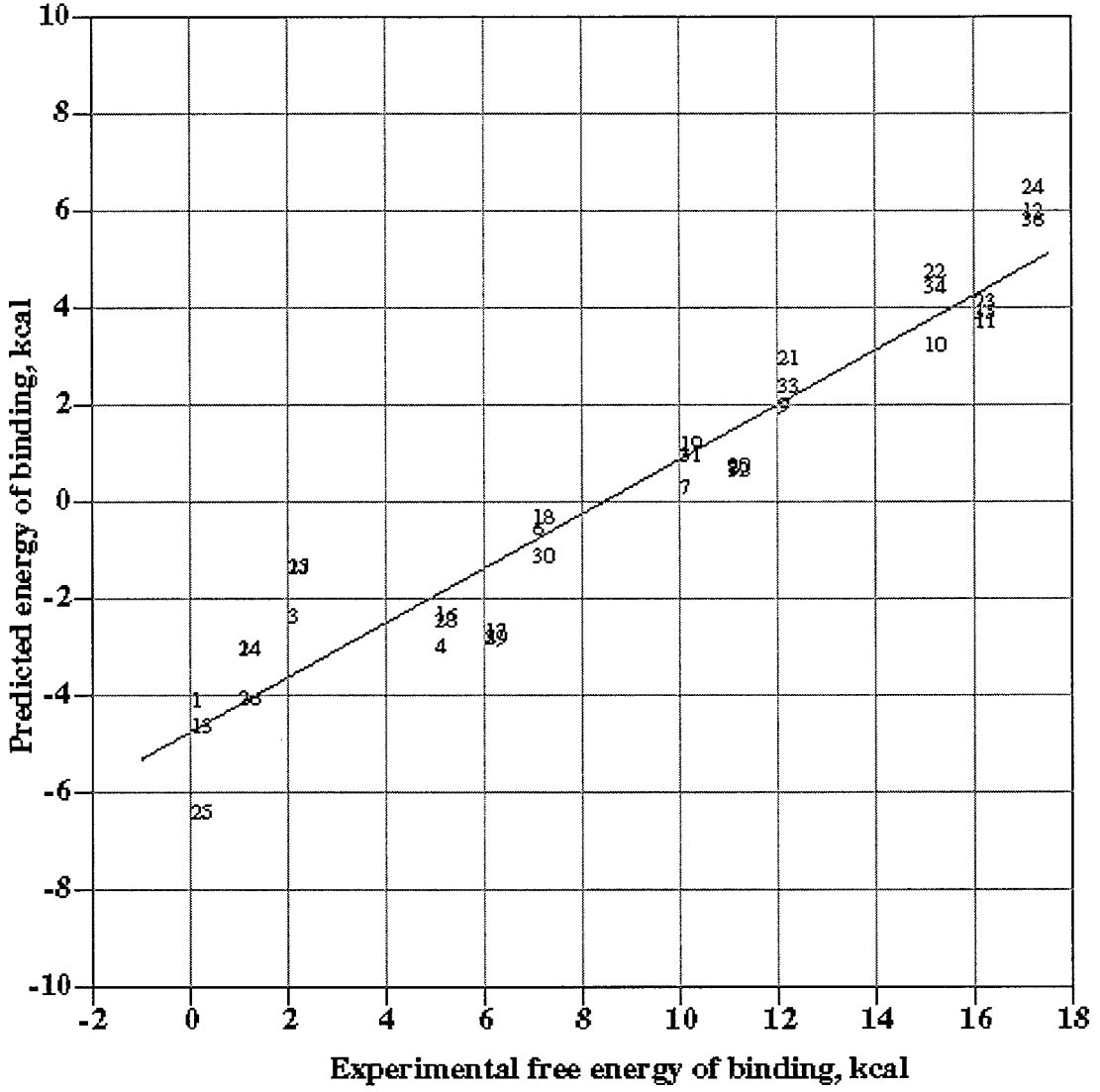


Figure 1. N-diaminoalkylpyrrole containing polyamide in the DNA minor groove. This globally optimized conformation shows interaction of the aminoalkyl tail with the DNA phosphates, which ensures inhibition of major-groove binding transcription factors by this type polyamides.



Eight-ring hairpins	
1	"AGTACT+IHPP_IHPP"
2	"AGATCT+IPHP_IPHP"
3	"AGTACT+IPPP_IPPP"
4	"AGTAt T+IPPP_iPPP"
5	"AGTATT+IPPP_PPPP"
6	"AGTAc T+IPPP_pPPP"
7	"AGGGAA+IIIP_PPPP"
8	"AGGc AA+IIIP_PPPP"
9	"AGTACT+IPiP_pPPP"
10	"TGGTCA+IIPP_IPPP"
11	"TGGTCA+IIPP_IhPP"
12	"TGGTCA+IIHP_IPPP"
13	"TGGACA+IIPP_IPPP"
14	"TGGACA+IIPP_IHPP"
15	"TGGACA+IIhP_IPPP"
16	"TGGGCA+IIhP_IpPP"
17	"TGGCCA+IIhP_IpPP"
18	"TGt TCA+IIPP_IPPP"
19	"TGGGt c+IIIP_PPPP"
20	"TGGGc t +IIIP_PPPP"
Six-ring hairpins	
21	"ATGTTA+IPP_PPP"
22	"ATGAc A+IPP_PPP"
23	"ATGGTT+IIP_PPP"
24	"ATGt TA+IIP_PPP"
25	"AgGGTA+IIP_PPP"
26	"AAACCA+PPP_IIP"
27	"ATAa CA+PPP_IIP"
28	"ATACCc +PPP_IIP"
29	"ATGTTT+IPP_PPP"
30	"ATGTTA+IPP_PPP"
31	"ATGTAA+IPP_PPP"
32	"ATGTAT+IPP_PPP"
33	"ATGATT+IPP_PPP"
34	"ATGATA+IPP_PPP"
35	"ATGAAA+IPP_PPP"
36	"ATGAAT+IPP_PPP"

Figure 2. Accuracy of binding energy predictions by ICFF modeling, tested on a set of available experimental data for short hairpins.

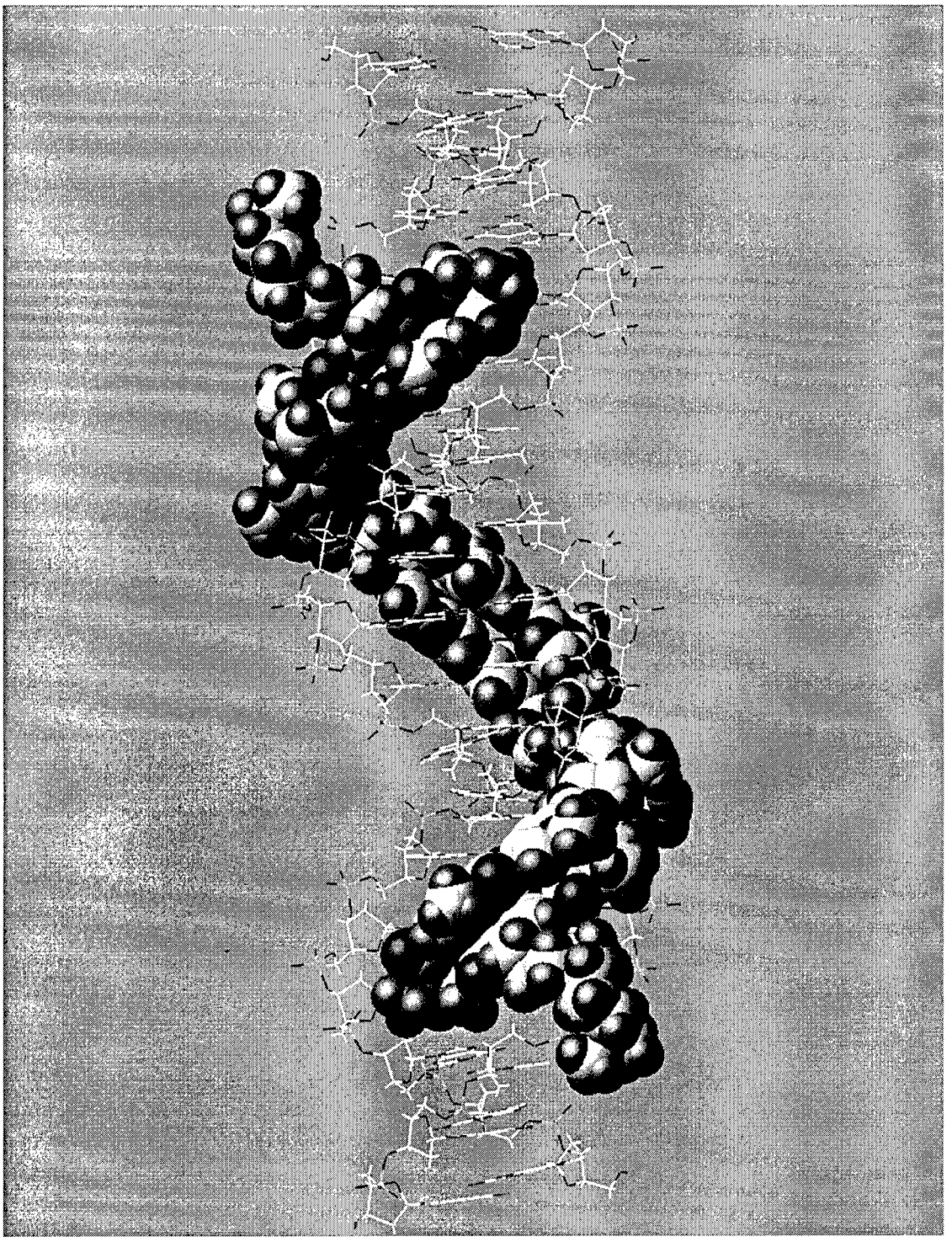


Figure 3. Recognition of a target DNA sequence **AGCGCGCTTGCT** by two sequence-specific polyamide hairpins, each containing 8 Im-Py rings. One of the pyrroles in each molecule is substituted by N-aminoalkylpyrrole.